DOCUMENT RESUME

ED 319 750 TM 014 836

AUTHOR Zeng, Lingjia; Bashaw, Wilbur L.
TITLE The Estimate of Difficulty and Dimensional Strengths of Test Items That Measure More Than One Ability.
PUB DATE Apr 90
NOTE 20p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Ability Identification; *Algorithms; Computer Simulation; *Difficulty Level; *Estimation (Mathematics); Item Response Theory; Maximum Likelihood Statistics; *Test Items
IDENTIFIERS Ability Estimates; Dimensional Analysis

ABSTRACT

A joint maximum likelihood estimation algorithm, based on the partial compensatory multidimensional logistic model (PCML) proposed by L. Zeng (1989), is presented. The algorithm simultaneously estimates item difficulty parameters, the strength of each dimension, and individuals' abilities on each of the dimensions involved in arriving at a correct response to the items. A maximum number of iterations was set for each estimation phase. In the model, iterations in a phase continue until either convergence is obtained or the maximum number of iterations is reached. In one particular phase, either item or person parameters are estimated, with the other parameter fixed. Estimation stops when the improvements are less than the prior chosen critical values. A parallel processing program, written in FORTRAN, was developed. This program and the algorithm were evaluated with several simulated item response data sets. Comparison of the estimates and the true simulated parameters suggested that the estimation was very accurate. Four tables present data from the simulations. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED319750

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
LINGJIA ZENG
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

The Estimate of Difficulty and Dimensional Strengths of Test Items that Measure More than One Ability

Lingjia Zeng and Wilbur L. Bashaw

The University of Georgia

Athens, Georgia 30602

April, 1990

TM014836

BEST COPY AVAILABLE

The Estimate of Difficulty and Dimensional Strengths of Test Items that Meassure More than One Ability

A number of authors have provided evidence that many test items require more than one ability for the examinees to obtain a correct response. However, the commonly used unidimensional item response theory (IRT) models, such as the Rasch model, two- or three-parameter models, are based on a strong assumption of a single trait underlying the individual's performance on the items. The studies concerning the robustness of the unidimensional IRT models suggest that applying these models to a clearly multidimensional data set is not appropriate. Evidently, in such cases, a model which can handle more than one ability dimension is more desirable.

McKinley and Reckase (1983), Reckase (1985) and Carlson (1988) reported studies on a mulitidimensional 2-parameter logist model (M2PL). The M2PL model includes a difficulty parameter, a vector of discrimination parameters of the item, and a vector of abilities of the individual who responds to the item. In terms of the probability of obtaining a correct response to the item, the M2PL model can be called a compensatory model, because it allows high ability on one dimension to compensate for low ability on the other dimension in arriving at a correct response to a test item. According to this compensatory model, the probability of getting a correct response to an item increases monotonically as the abilities involved in the item increase. Several authors have argued that the compensatory model is not very realistic in a multidimensional testing situation where the items are considered as requiring simultanous application of more than one ability to get a correct response (Ansley and Forsyth, 1985; Way, Ansley and Forsyth, 1988). For example, if a math problem solving item requires both reading and computation abilities, a student who cannot read the item will certainly not be able to answer it correctly, regardless of

correctly, regardless of how well he may perform conputation. But according to the compensatoy model, if a student has a very high computation ability he will have a very good chance to answer a math problem-solving item correctly even if he cannot read the item.

The opposite of compensatory model is a noncompensatory model, such as a model proposed by Sympson (1978) (note, Sympson originally called it a partial noncompensatory model). The noncompensatory model does not allow any compensation. For example, if a test item requires two abilities, a low ability on either dimension will yield a low probability of obtaining a correct response. Sympson's model has been used by a number of authors to generate artificial item response data for the purpose of studying the robustness of unidimensional IRT models (e.g. Ansley and Forsyth, 1985; Hambleton and Rovinelli, 1986). However, a parameter estimation algorithm based on a noncompensatory multidimensional model still remains a unsolved problem.

Both the compensatory or noncompensatory model focus on the number of dimensions involved in the test, but ignore the different strength each dimension may have in determining the probability of getting a correct response. For example, a math problem-solving item may require reading, computation and logical reasoning abilities. Testing reading ability, however, may not be the goal of such an item, and the item may be constructed with a very low requirement of reading ability, while the computation or logical reasoning may be the primary abilities the item is designed to measure. In terms of the probability of arriving at a correct response for a student, his reading ability accounts for only a small proportion of the total performance, while the computation and logical reasoning abilities account for a majority proportion.

Zeng (1989) proposed a partial compensatory multidimensional logistic model (PCML), which is a modification of Sympson's noncompensatory model. PCML model has one more parameter than the noncompensatory model, which is the dimensional strength parameter *w*. The *w* parameter defines the proportion an ability dimension accounts for in to the total performance on an item. The PCML

model was first used to generate artificial response data for the purpose of studying the robustness of unidimensional IRT models (Zeng, 1989). This paper introduces a joint maximum likelihood estimation algorithm based on the PCML model, which simultaneously estimates item difficulty parameters, the strength of each dimension, and individuals' abilities on each of the dimensions involved in arriving at a correct response to the items.

The Model

The PCML model used in this study is defined by the following equation:

$$P_{ij}\,(x_{ij} = 1 \mid b_i, \mathbf{w}_i, \boldsymbol{\theta}_j) = \prod_{k=1}^{K} s_{ijk}^{\,w_{ik}} \qquad [1]$$

$$i = 1, 2 \ldots M,\ j = 1, 2 \ldots N,\ k = 1, 2 \ldots K$$

where:

P_{ij} = the probability of a correct response to item i by individual j;

x_{ij} = the response of individual j on item i (1 = correct, 0 = incorrect);

b_i = the difficulty parameter of item i;

$\mathbf{w}_i$ = a vector of K dimensional strength parameters of item i;

$\boldsymbol{\theta}_j$ = a vector of K ability parameters of individual j;

$$s_{ijk} = \frac{1}{1 + \exp\left[-D\left(\theta_{jk} - b_i\right)\right]} \qquad [2]$$

where D is an arbitrary scaling factor. s_{ijk} is the probability of performing correctly on dimension k of item i, which is determined by the individual's ability on

dimension k, and b_i, the item difficulty. The proportion of total performance which s_{ijk} accounts for is determined by w_{ik}, the dimensional strength parameter. The relationship between the dimensional strength and the probability of a correct item response is illustrated in an example given in Table 1. Suppose there is a set of 7 items which require a student to simultaneously apply 2 abilities to answer each item correctly, and the student has a low ability on dimension 1 and high ability on dimension 2. Let's further suppose that all the items have the same difficulties, so the probability of performing correctly on the two dimensions will be the same across item 1 through item 9, say, 0.2 for s_1 and 0.9 for s_2. For item 1, $w_1 = 0$, which means that this item does not require any ability on dimension 1 at all, so P, the probability of a correct response to item 1 is equal to 0.9 which is the value of s_2. It is clear as shown in Table 1, that when a dimensional strength is small, this dimension will have little to contribute to P, and that the value of P is closer to the value of s which has a greater dimensional strength w. Equation 2 is in fact a unidimensional one-parameter IRT model. If there is a dimension which accounts for approximately 100 percent of the total performance, the probability value obtained from equation 1 will be the same as the one obtained from equation 2 (e.g. item 1 and item 7 in Table 1.

---

insert Table 1. here

The probabilities of a correct response calculated with compensatory and noncompensatory models are also listed in Table 1. The compensatory model yields a probability of correct response of 0.69 for all of the seven items. Even for items 6 and 7, in which dimension1 accounts for more than 90 percent of the total performance, and the student's probability of doing correctly on this dimension is as low as 0.2, he still has a very good chance (0.69) to answer the item correctly, according to the compensatory model. The noncompensatory model yields a probability of correct response of 0.18 for all the items. The value is obviously not reasonable for this student's performance on item 1 and 2. Although he has a very

low ability on dimension 1, he still should have a very good chance to answer these 2 items correctly, because dimension 1 only accounts for a very small proportion of the total performance on these two items, and the probability of his doing correctly on the major dimension is 0.90.

The Joint Maximum Likelihood Estimation Algorithm

In order to derive a likelihood function to estimate the parameters, it is necessary to assume that the individual's response to each test item, conditional to his ability vector, is independent of every other item, and each individual's responses to the items are independent of those of others. Hence, the likelihood of a set of N individuals' responses to M items is a product of probabilities:

$$L = \prod_{j=1}^{N} \prod_{i=1}^{M} P_{ij}^{x_{ij}} \quad Q_{ij}^{(1-x_{ij})} \qquad [3]$$

where $Q_{ij} = (1 - P_{ij})$

The values which maximize the likelihood function L are the maximum likelihood estimates of the parameters of the model. Since the maximum occurs for the same values in the logarithm of L as in L, we can maximize the natural logarithm of L, which makes the computation much easier. The natural logarithm of L is:

$$\ln(L) = \sum_{j=1}^{N} \sum_{i=1}^{M} [\, x_{ij} \ln(P_{ij}) + (1 - x_{ij}) \ln(Q_{ij})] \qquad [4]$$

Since maximizing $\ln(L)$ is the same as minimizing the negative of $\ln(L)$, we are going to minimize $-\ln(L)$. Thus, we define the maximum likelihood estimates of the parameters of the model as the following values:

b_i, w_i, θ_j $(i = 1,2...M; j = 1,2,...N)$, which minimize

$$f(y) = - \ln(l) \qquad [5]$$

where y is a vector of parameters to be estimated. Thus, the problem is reduced to a nonlinear system of equations:

$$g(y) = 0 \qquad [6]$$

where $g(y)$ is a vector of the first partial derivatives of f, $(\partial f/\partial b, \partial f/\partial w_1, \partial f/\partial w_2....)$, $g(y)$ is also called the gradient of f. The equation system (6) can be solved by an iterative procedure such as the Newton-Raphson method. Starting at some initial estimates $\hat{y}_k$, some better estimates $\hat{y}_{k+1}$ can be obtained by:

$$\hat{y}_{k+1} = \hat{y}_k - H(\hat{y}_k)\, g(\hat{y}_k) \qquad [7]$$

where $H(\hat{y})$ is the inverse of the Jacobian matrix $J(\hat{y})$ of g. $J(\hat{y})$ is also called the Hessian, the matrix of second partial derivatives of f.

$$H = J^{-1} \qquad [8]$$

The joint ML estimation procedure as usually applied to the unidimensional IRT model parameter estimation (e.g. Wingersky, Barton, and Lord, 1982) can be adopted to fit the multidimensional parameter estimation. In the present study, we set a maximum number of iterations for each estimation phase. Iterations in a phase continue until either convergence is obtained of the maximum number of iterations is reached. In one particular phase, either item or person parameters are estimated. When the iterations are performed for estimating the item parameters, the person parameters are fixed at their current values, and vice versa. The estimation procedure stops when the improvements are less than the prior chosen critical values.

Note that **H** and **J** are symmetric matrices. The upper diagonal entries of the Jacobian matrix for estimating item parameters are:

$$\mathbf{J}(\hat{y}) = \begin{bmatrix} \frac{\partial^2 f}{\partial^2 b_i} & \frac{\partial^2 f}{\partial b_i \, \partial w_{i1}} & \frac{\partial^2 f}{\partial b_i \, \partial w_{i2}} & \cdots \\ & \frac{\partial^2 f}{\partial^2 w_{i1}} & \frac{\partial^2 f}{\partial w_{i1} \, \partial w_{i2}} & \cdots \\ & & \frac{\partial^2 f}{\partial^2 w_{i1}} & \cdots \\ & & & \cdots \end{bmatrix} \qquad [9]$$

And the upper diagonal entries of the Jacobian matrix for estimating the person parameters are:

$$\mathbf{J}(\hat{y}) = \begin{bmatrix} \frac{\partial^2 f}{\partial^2 \theta_{j1}} & \frac{\partial^2 f}{\partial \theta_{j1} \, \partial \theta_{j2}} & \cdots \\ & \frac{\partial^2 f}{\partial^2 \theta_{j2}} & \cdots \\ & & \cdots \end{bmatrix} \qquad [10]$$

The analytic expressions for the first partial derivatives are derived as follows:

$$\frac{\partial f}{\partial b_i} = \sum_{j=1}^{N} \frac{\partial P_{ij}}{\partial b_i} \frac{P_{ij} - x_{ij}}{P_{ij}\ Q_{ij}} \qquad [11]$$

where $\frac{\partial P_{ij}}{\partial b_i} = D \sum_{k=1}^{K} -w_{ik}\ t_k\ (1 + t_k)^{-w_{ik}-1} \prod_{\substack{m=1 \\ m \neq k}}^{K} (1 + t_m)^{-w_{im}}$

where $t_k = \exp\left(D(b_i - \theta_{ik})\right)$

$$\frac{\partial f}{\partial w_{ik}} = \sum_{j=1}^{N} \frac{\partial P_{ij}}{\partial w_{ik}} \frac{P_{ij} - x_{ij}}{P_{ij}\ Q_{ij}} \qquad [12]$$

where $\frac{\partial P_{ij}}{\partial w_{ik}} = \ln(s_{ijk}) \prod_{m=1}^{K} s_{ijm}^{\ w_{im}}$

$$\frac{\partial f}{\partial \theta_{jk}} = \sum_{i=1}^{M} \frac{\partial P_{ij}}{\partial \theta_{jk}} \frac{P_{ij} - x_{ij}}{P_{ij}\ Q_{ij}} \qquad [13]$$

where $\frac{\partial P_{ij}}{\partial \theta_{jk}} = D\ w_{ik}\ t_k\ (1 + t_k)^{-w_{ik}-1} \prod_{\substack{m=1 \\ m \neq k}}^{K} s_{ijm}^{\ w_{im}}$

The Jacobian matrices used in this study were not calculated directly with analytic expressions of the second partial derivatives but approximated from the first partial derivatives by using Taylor's polynomials. The detail of this approach can be found in any numerical analysis textbood (e.g., Burden and Faires, 1985).

Because of the indeterminant nature of the model, it is neccessary to rescale the estimates of the parameters after several phases of iterations. In this study, the

estimates of θ and b were rescaled to zero mean and unit variance after each phase in which they were estimated. We found that the rescaling of θ and b resulted in a very small change in the function value of $s_{\cdots ijk}$, so, the estimates of w were not rescaled accordingly. It is also necessary to constrain the estimates of the parameters to prevent them from drifting to unreasonably large or small values. The b and θ were constrained in -4 to 4, w was constrained in 0 to 1.0.

The standard errors of the estimated parameters were computed by taking the square roots of the diagonal entries of the inverse of the Jacobian matrices, the matrices of second derivatives, evaluated at the last iteration. As pointed out by Hambleton and Swaminathan (1985), these standard errors are only approximations of the true values.

Computer Program

Solving the nonlinear system 6 involves excessive computations. The most expensive part of the algorithm is to find the Jacobian matrices. The dimension of the Jacobian matrix for estimation item parameters is $O(K^2)$ (where K is the number of dimensions). The computation of each entry of J, involves MK function evaluations (where M is the number of items), and this procedure needs to be repeated N times (N is the number of persons). Therefore, the time complexity of this algorithm is $O(MNK^3)$, which means that the computation time increases cubically as the number of dimensions increases. The conventional computer, including main frame computer, will be inadquate when the dimension exceeds five. To solve this problem, a supercomputer, a computer with multiprocessors, will be needed.

In this study, we developed a parallel processing program written in FORTRAN. This program was implemented on a Hypercube system, which can be considered as a "super microcomputer". This Hypercube had 16 80386 processors, eight of which had vector processing boards. This program utilized the eight processors with vector boards.

Evaluation of the Estimation Program

A number of simulated item response data sets were generated to evaluate the performance parameter estimation program. The simulated data were generated in the following way. The item difficulty parameters b were randomly selected from a uniform distribution on the interval between -2.0 and 2.0. The dimension strength parameters w_{i1} were randomly drawn from a uniform distribution on the interval (0, 1), and w_{i2} were set to 1 - w_{i1}. The person ability parameters, θ, were randomly selected from a uniform distribution on the interval (-2.0, 2.0). The inter-dimension correlation between the two vectors was less than 0.01. For each simulated subject j on each simulated item i, the probability of arriving at a correct response P_{ij} was computed using equation 1, and then P_{ij} was compared to a random variable R which was drawn from a uniform distribution on the interval (0, 1). The simulated response score x_{ij} was assigned a value 1 for a correct response if $P > R$, otherwise, 0 was assigned for an incorrect response. This resulted in an NxM matrix of response scores with either 1 or 0 as their entries.

Joint maximum likelihood estimation was performed on these simulated item response score data sets. The item and person parameters estimated from the simulated item response data sets were compared with the true (simulated) parameters. The correlation coefficients between the true and the estimated parameters were calculated and the results are presented in Table 2.

---

insert table 2 here

In general, the correlation between the true and the estimated item and person parameters are all quite good for all the data sets as it is shown in Table 2. The correlation between the true and estimated difficulty parameter, b, and dimension strength parameter, w, are very high even with a sample size as low as 200 subjects by 24 items. This result suggests that item parameter estimation is very accurate.

The corrlations between the true and estimated person ability parameter turned out to be smaller than those between the true and estimated item parameters. The

quality of parameter estimation tends to related to the length of the test. From Table 2 one can see that the parameters estimated from 40-item data sets are generally better than those from 24-item data sets, and the parameter obtained from 80-item data sets are generally better than those from 40-item data sets. This trend is reasonable, because a longer test should yield better estimates of the examinee's abilities than a shorter test, if other factors such as time, fatigue, etc., are not considered. When the test length dropped to 24 items, the estimated θ parameters had very low correlations with the true parameters ($r<0.50$).

A sample of the true and estimated difficulty parameters, and the standard errors of the estimates are listed in Table 3. These values were calculated from a data set of 200 persons by 40 items. The true and estimated dimensional strength parameters and the standard errors of the estimates, calculated from the same data set, are listed in Table 4. Is can be seen from Table 3 that the difficulty parameters estimation was very accurate for all items, with very small standard errors.

---

insert Table 3 here

The w parameter estimations were accurate enough for most items. But for a few items, such as items 7, 8, 12, 13, 15 and 31, the w parameter estimations were not accurate. The standard error of the estimates were quite large, greater than 0.5 (see table 4). The estimated values for these items exceeded the range of 0 to 1.0, and the program assigned them the extreme values.

---

insert table 4 here

Possible Applications

We have shown that a joint maximum likelihood estimation algorithm is feasible to simultaneously estimate from a set of item response scores all the parameters of the multidimensional model, that is, the item difficulty, dimensional strengths and

the examinees' abilities on each of the dimensions involved in arriving at a correct response to an item. A possible application is to use this model to develop a multidimensional instrument which consists of items each of which measures one or more abilities. Analyzing item response scores with this model, each item can contribute more or less information about an examinee's abilities on several dimensions, which can not be obtained from a unidimensional model.

In most test situations, we may use items from an item bank, in which all the items were caliberated by applying the multidimensional model to a very large data set. In such cases we could assume we know all the difficulty and dimensional strength parameters of the items. In this situation, estimation of an examinee's abilities can be performed on his responses to a set of items. On the other hand, if we know a group of examinees' abilities from some previous tests, the item difficulty and dimensional strength parameters can be esimated from these examinees' responses to a single item. In either cases, the parameter estimation is much easier and could be more accurate than simultaneous estimation of all parameters.

Table 1. An Example Showing the Relationship Between the Dimensional Strength and the Total Performance

Item	w_1	w_2	s_1	s_2	P_1	P_2	P_3
1	0.0	1.0	0.20	0.90	0.90	0.69	0.18
2	0.1	0.9	0.20	0.90	0.77	0.69	0.18
3	0.3	0.7	0.20	0.90	0.57	0.69	0.18
4	0.5	0.5	0.20	0.90	0.42	0.69	0.18
5	0.7	0.3	0.20	0.90	0.32	0.69	0.18
6	0.9	0.1	0.20	0.90	0.23	0.69	0.18
7	1.0	0.0	0.20	0.90	0.20	0.69	0.18

Note:
1 for partial compensatory model, 2 for compensatory model and 3 for noncompensatory model.

$$P_2 = \frac{1}{1 + \exp\left[-D \sum_{k=1}^{2} y_k\right]}$$

$$P_3 = \frac{1}{\prod_{k=1}^{2} \left[1 + \exp\left[-D\, y_k\right]\right]}$$

where $D = 1.7$, and $y_1 = -0.815$, and $y_2 = 1.292$

Table 2 The Correlations between the True and Estimated Parameters

Sample Size	Correlations $b\hat{b}$	$w\hat{w}$	$w\hat{w}$	$\theta\hat{\theta}$	$\theta\hat{\theta}$
400 80	0.9967	0.9392	0.9443	0.7006	0.7275
200 80	0.9910	0.9430	0.9425	0.8461	0.7280
400 40	0.9947	0.9064	0.9086	0.7610	0.6091
200 40	0.9923	0.9238	0.9263	0.6997	0.6317
400 24	0.9926	0.8327	0.8217	0.4472	0.4630
200 24	0.9886	0.8261	0.8210	0.4948	0.4963

Table 3 The True and Estimated b Parameters and the Standard Errors of Estimates

Item	b	$\hat{b}$	SE
1	0.414	0.253	0.101
2	0.587	0.691	0.115
3	0.725	0.623	0.108
4	1.938	1.944	0.198
5	1.363	1.110	0.135
6	-0.764	-0.612	0.100
7	-1.599	-1.501	0.132
8	1.364	1.394	0.152
9	-0.417	-0.398	0.099
10	0.920	0.807	0.120
11	0.826	0.576	0.110
12	0.806	0.886	0.121
13	0.380	0.238	0.101
14	-1.768	-1.541	0.132
15	-1.386	-1.217	0.117
16	0.150	0.129	0.101
17	-1.152	-1.027	0.107
18	-1.931	-1.679	0.139
19	-0.403	-0.206	0.098
20	1.150	0.837	0.119
21	0.781	0.677	0.113
22	0.649	0.613	0.112
23	-0.574	-0.367	0.099
24	-0.414	-0.484	0.100
25	0.163	0.118	0.102
26	0.691	0.727	0.111
27	0.528	0.452	0.104
28	1.970	1.844	0.190
29	0.909	0.937	0.124
30	-0.368	-0.247	0.098
31	-1.511	-1.339	0.123
32	0.957	0.989	0.126
33	-1.566	-1.281	0.117
34	-0.840	-0.489	0.098
35	-1.460	-1.214	0.120
36	0.166	0.153	0.100
37	-1.099	-0.755	0.103
38	1.085	1.047	0.132
39	0.799	1.097	0.132
40	-0.876	-0.757	0.104

Table 4 The true and Estimated w Parameters and the Standard Errors of Estimates

item	w_1	w_2	$\hat{w}_1$	$\hat{w}_2$	SE_2	SE_2
1	0.840	0.160	0.999	0.001	0.225	0.070
2	0.447	0.553	0.247	0.685	0.096	0.122
3	0.886	0.114	0.999	0.001	0.203	0.090
4	0.584	0.416	0.763	0.216	0.159	0.109
5	0.384	0.616	0.474	0.498	0.130	0.132
6	0.209	0.791	0.001	0.999	0.630	0.614
7	0.164	0.836	0.001	0.999	0.884	0.856
8	0.533	0.467	0.676	0.308	0.155	0.124
9	0.583	0.417	0.658	0.334	0.137	0.109
10	0.390	0.610	0.349	0.614	0.113	0.130
11	0.742	0.258	0.725	0.259	0.140	0.105
12	0.226	0.774	0.001	0.999	0.502	0.500
13	0.857	0.143	0.999	0.001	0.517	0.517
14	0.663	0.337	0.582	0.450	0.189	0.172
15	0.153	0.847	0.001	0.999	0.780	0.744
16	0.533	0.467	0.237	0.775	0.114	0.145
17	0.266	0.734	0.190	0.877	0.098	0.164
18	0.307	0.693	0.423	0.636	0.186	0.215
19	0.609	0.391	0.773	0.237	0.136	0.094
20	0.886	0.114	0.698	0.275	0.130	0.100
21	0.152	0.848	0.001	0.999	0.501	0.499
22	0.387	0.613	0.375	0.595	0.108	0.121
23	0.754	0.246	0.999	0.001	0.206	0.062
24	0.532	0.468	0.721	0.269	0.142	0.104
25	0.652	0.348	0.520	0.414	0.115	0.108
26	0.946	0.054	0.999	0.001	0.436	0.178
27	0.944	0.056	0.999	0.001	0.510	0.509
28	0.517	0.483	0.592	0.366	0.165	0.148
29	0.174	0.826	0.001	0.999	0.043	0.133
30	0.205	0.795	0.001	0.999	0.222	0.348
31	0.209	0.791	0.001	0.999	0.758	0.751
32	0.303	0.697	0.001	0.999	0.125	0.271
33	0.498	0.502	0.416	0.641	0.156	0.180
34	0.286	0.714	0.198	0.824	0.092	0.139
35	0.676	0.324	0.999	0.001	0.274	0.081
36	0.810	0.190	0.999	0.001	0.204	0.080
37	0.377	0.623	0.001	0.999	0.068	0.220
38	0.355	0.645	0.211	0.796	0.100	0.140
39	0.670	0.330	0.655	0.287	0.134	0.103
40	0.566	0.434	0.719	0.265	0.156	0.113

REFERENCES

Ansley, T. N., & Forsyth, R. A. (1985). Unidimensional estimates from two-dimensional data. Applied Psychological Measurement, 9, 37-47.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M . R. Novick (Eds.), Statistical theories of mental test scores. (pp. 397-472). Reading, MA: Addison-Wesley.

Burden, R. L. & Faires, J. D. (1985). Numerical analysis. Boston, MA: PWS-KENT Publishing Company.

Carlson, J. E. (1987). Multidimensional item response theory estimation: a computer program. (Research Report ONR 87-2). Iowa City, IA: The American College Testing Program.

Hambleton, R. K., & Rovinelli, R. J. (1986). Assessing the dimensionality of a set of test items. Applied Psychological Measurement, 10, 287-302.

Hambleton, R. K. & Swaminathan, H. (1985). Item response theory: principles and applications. Boston, MA: Kluwer Nijhoff.

Lord, F. M. (1980). Application of item response theory to practical testing problems. Erlbaum, NJ: Hillsdale.

McKinley, R. L., & Reckase, M. D. (1983). An extension of the two-parameter model to the multidimensional latent space. (Research Report ONR 83-2). Iowa City, IA: The American College Testing Program.

Reckase, M. D., (1985). The difficulty of test items that measure more than one ability. Applied Psychological Measurement, 9, 401-412.

Sympson, J. B. (1978). A model for testing with multidimensional items. In D.J. Weiss (Ed.) Proceedings of the 1977 Computerized Adaptive Testing Conference, Minneapolis, MN: Psychometric Lab.

Way, A. D.; Ansley, T. N.; & Forsyth, R. A. (1988). The comparative effects of compensatory and noncompensatory two-dimensional data on unidimensional IRT estimates. Applied Psychological Measurement, 12, 239-252.

Wingersky, M. S., Barton, M. A., & Lord, F. M. (1982). LOGIST User's Guide. Princeton, NJ: Educational Testing Service.

Zeng, L. (1989). Robustness of unidimensional latent trait models when applied to multidimensional data. Unpublished doctoral dissertation, Athens, GA: University of Georgia.